CLAIMS

1. A computer-implemented method for predicting at least one amino acid sequence compatible with a specified three-dimensional (3D) structure of a protein or peptide, which method comprises the steps of:-

(a) providing a coordinate set representing the backbone of said 3D structure;

(b) constructing a reduced virtual representation for the 3D structure provided in step (a);

(c) determining for each position along the virtual structure representation provided in step (b) its solvent accessibility;

(d) constructing an initial amino acid sequence by randomly assigning for each position along the structure an amino acid residue selected randomly from a predefined group of amino acids having a solvent accessibility compatible with the solvent accessibility of said position;

(e) randomly selecting one or more positions along the sequence provided in step (d) and applying on each position a Monte-Carlo simulation in sequence space and rotamer space, said simulation comprising one or more scoring function calculating steps which include:-

i) randomly selecting one or more amino acid residues of the same solvent accessibility as that defined for said position to obtain a mutation;

ii) calculating an energy scoring function for each possible rotamer of each amino acid residue provided in step (i) based on their said reduced virtual representation;

iii) selecting the lowest scoring rotamer or when more than one amino acid is manipulated simultaneously, selecting the lowest

scoring rotamer combination;

iv) determining whether to accept or reject the mutation with the rotamer or rotamer combination selected in step (iii); and

v) assigning the amino acid residue or residues and their respective selected rotamer or rotamer combinations selected in step (iii) to said position/s and moving to another position along the sequence;

said simulation steps are repeated until for each position along said sequence, the residue and residue's rotamer with the lowest energy score is selected, to obtain a virtually represented amino acid sequence with the lowest total energy score;

(f) expanding the reduced representation of the virtually represented amino acid sequence obtained in step (e) to its corresponding all-atom sequence representation thereby obtaining an amino acid sequence compatible with the predefined 3D structure.

(g) optionally, creating a computer output of the expanded all-atom representation of the primary structure/s obtained in step (f).

2. The method as claimed in claim 1, wherein the 3D structure provided in step (a) is that of a native peptide, or protein, or of a designed protein.

3. The method as claimed in claim 1, wherein said coordinate set is provided in a computer readable form.

4. The method as claimed in claim 1, wherein said amino acid sequence may comprise naturally occurring amino acid residues, synthetic amino acid residues, or variations of said naturally occurring or synthetic amino acid residues.

5. The method as claimed in claim 1, wherein for each position along the 3D structure its solvent accessibility is determined according to the extent of exposure of said position to the solvent surrounding it, said position being either buried, exposed or

intermediate position.

6. The method as claimed in claim 5, wherein said solvent is substantially water.

7. The method as claimed in claim 6, wherein said buried positions are occupied by hydrophobic amino acid residues.

8. The method as claimed in claim 7, wherein said hydrophobic amino acid residues are selected from the group consisting of Ala, Tyr, Trp, Val, Leu, Ile, Phe, Met, Cys, Pro, Gly.

9. The method as claimed in claim 5, wherein said exposed positions are occupied by hydrophilic amino acid residues.

10. The method as claimed in claim 9, wherein said hydrophilic amino acid residues are selected from the group consisting of Lys, Arg, His, Glu, Asp, Gln, Asn, Ser, Thr.

11. The method as claimed in claim 5, wherein said intermediate positions are occupied by either hydrophilic or hydrophobic amino acid residues.

12. The method as claimed in claim 11, wherein said intermediate positions are occupied by amino acid residues selected from the group consisting of Pro, Lys, Arg, His, Glu, Asp, Gln, Asn, Ser, Thr, Gly, Ala, Tyr, Trp, Val, Leu, Ile, Phe, Met, Cys.

13. The method as claimed in claim 1, wherein said Monte Carlo simulation is applied simultaneously on up to three random positions in said sequence.

14. The method as claimed in claim 1, wherein said Monte Carlo step is conducted either at a fixed temperature or at a varying annealing temperature.

15. The method as claimed in claim 1, wherein a *de novo* amino acid sequence is generated.

16. The method as claimed in claim 1, wherein said amino acid sequence folds under physiological condition into a biologically functional 3D conformation substantially identical to said predefined 3D structure or to a portion thereof.

A

17. The method as claimed in claim 15 ~~or 16~~, wherein said *de novo* amino acid sequence stabilized said 3D structure, as compared to the native amino acid sequence.

18. An amino acid sequence which folds under physiological conditions into a specified 3D structure, said amino acid sequence is obtained by the method of claim 1.

19. An amino acid sequence according to claim 18, which is biologically functional.

A

20. A nucleic acid sequence encoding the amino acid sequence of claim ~~19 or 20~~ 18.

21. A computer-based system for predicting an amino acid sequence compatible with a predefined 3D structure according to the method of claim 1, said system comprising:-

(a) input apparatus for specifying said 3D structure;

(b) a first memory for storing the specified 3D structure;

(c) a second memory having a stored thereon an application program which when running, provides at least one amino acid sequence compatible with the specified 3D structure;

(d) a third memory for storing the at least one amino acid sequence obtained;

(e) a processor coupled to said input means, and to said first, second and third memories for representation of said amino acid sequence; and

(f) optionally, a display unit coupled to said processing means for displaying the amino acid sequence.

Cont A1